ED 330 725 TM 016 332

AUTHOR Friedrich, Katherine R.
TITLE The Importance of Structure Coefficients in Parametric Analyses.
PUB DATE Apr 91
NOTE 18p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, April 3-7, 1991).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Analysis of Variance; *Factor Analysis; *Heuristics; *Multiple Regression Analysis; *Predictor Variables; Programing
IDENTIFIERS Collinearity; *Parametric Analysis; *Structure Coefficients; Univariate Analysis

ABSTRACT

The recognition that all parametric methods are interrelated, coupled with the notion that structure coefficients are often vital in factor and canonical analyses, suggests that structure coefficients may be important in univariate analysis as well. Using a small, heuristic data set, this paper discusses the importance of structure coefficients for the univariate case and then extends this discussion to all parametric procedures. More specifically, a multiple regression analysis, with three independent variables, was conducted. Examination of the intercorrelations between predictors indicated that there was some degree of collinearity. Beta weights, structure coefficients, and other parametric procedures were used when appropriate. Regression coefficients and structure coefficients represent two different types of information. It is suggested that serious misinterpretations of results can occur if structure coefficients are not used in some analyses. The Statistical Analysis System program used to analyze the heuristic data involved in this study, three tables, and a 19-item list of references are included. (TJH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED330725

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
KATHERINE R. FRIEDRICH
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

The Importance of Structure Coefficients in Parametric Analyses

Katherine R. Friedrich

Texas A&M University 77843-4225

BEST COPY AVAILABLE

Paper presented at the annual meeting of the American Educational Research Association, Chicago, April 3-6, 1991.

TM016332

ABSTRACT

Two seminal articles (Cohen, 1968; Knapp, 1978) stimulated recognition that all parametric procedures can be derived from the same linear model involving least squares weights. The recognition that all parametric methods are interrelated, coupled with the feeling that structure coefficients [illegible] analyses such as factor analysis (Gorsuch, 1983) and canonical analysis (Thompson, 1984), stimulated recognition that structure coefficients may be important in univariate analysis as well (Thompson & Borrello, 1985). Using a small heuristic data set, this paper discusses the importance of structure coefficients for the univariate case and then extends this discussion to all parametric procedures. It is suggested that serious result misinterpretation can occur if structure coefficients are not utilized in some analyses.

The Importance of Structure Coefficients in Parametric Analyses

The concept of a single general linear model is not a new one. It was first introduced by Cohen (1968) in his discussion of multiple regression as a general analytic procedure. This notion was extended by other researchers (Baggaley, 1981; Fornell, 1978) who recognized that the most general case of the linear model is canonical correlation analysis, not multiple regression. The idea that all parametric methods involve least squares weights was clearly demonstrated in a seminal article by Knapp (1978), who noted that "virtually all of the commonly encountered parametric tests of significance can be treated as special cases of canonical correlation analysis, which is the general procedure for investigating differences between two sets of variables" (p. 410).

Given that all parametric methods are in actuality "special cases" of a single procedure, coupled with the value of structure coefficients in the interpretation of canonical correlation analysis (Thompson, 1984, 1988), it logically follows that structure coefficients may be utilized in all parametric procedures. The importance of structure coefficients to factor analysis was stressed by Gorsuch (1983) in his text on the procedure. Their usefulness in interpreting the results of a multivariate analysis of variance analysis was discussed by Daniel (1990), and other researchers (Perry, 1990; Thompson & Borrello, 1985) have demonstrated how applicable structure coefficients are to univariate analyses.

Even though it is now recognized fact that structure coefficients may calculated for any parametric analysis, they are not routinely reported and utilized in the interpretation of such analyses. This may result from the fact that this particular type of coefficient is more commonly associated with more complicated procedures such as canonical correlation and factor analysis. Kerlinger and Pedhazur (1973) noted that the mathematical sophistication required to carry out canonical correlation prohibited its use before the availability of computers. Since then, however, many authors have found that in addition to the mathematical knowledge required, a lack of familiarity and difficulty in interpretation has

unfortunately continued to inhibit researchers from employing this technique (Baggely, 1981; Kerlinger & Pedhazur, 1973; Thompson, 1980). Since structure coefficients are typically associated with canonical correlation and other complicated procedures such as factor analysis and discriminant analysis, it is not unfounded to conclude that the reluctance of researchers to employ such coefficients may stem from a general lack of understanding of this type of coefficient. The intent of this paper is to acquaint the reader with the concept of a structure coefficient in general terms and the importance of such coefficients to all parametric procedures. The approach taken here shall be to first define and illustrate structure coefficients using a small heuristic data set within a familiar context, the univariate case, and then to extend these ideas to all parametric methods.

The Univariate Case - Multiple Regression

The parametric procedure multiple regression is a useful context within which to discuss structure coefficients, for several reasons. As a statistical procedure, except for analysis of variance, multiple regression is more widely known and applied than other parametric procedures (Willson, 1980). Also, compared to multivariate methods, such as factor analysis or discriminant analysis, multiple regression is mathematically simpler, thus making the example more accessible. Next, because it is a univariate procedure, the interpretation of the results of a regression analysis is more straightforward. And finally, the basic mathematical model implicit in all parametric procedures is explicitly stated in multiple regression, thus enabling the reader to clearly see from where the components in this discussion are drawn. Although all analyses involve models, the models tested are not always explicit to users.

All parametric procedures are designed to investigate the relationship between two or more sets of variables where the researcher is primarily interested in predicting or explaining some phenomenon. In terms of multiple regression, the objective is to develop an equation where the predictor variables, denoted X_i, are optimally weighted such that the

distance (i.e., the e or residual score) between each individual's predicted score, Y*, and the individual's actual score, Y, is minimized. This is accomplished through a method of estimation known as least squares. A regression equation typically takes the following form:

$$Y \text{ <---- } Y^* = b_0 + b_1(X_1) + b_2(X_2)$$

where b_0, sometimes referred to as the Y-intercept, is the value of Y* when both predictor variables are equal to zero, and b_1 and b_2 are the regression coefficients or b weights. The weights (b_0, b_1, and b_2) are among the parameters estimated by the least squares method. The methods are called "least squares" because the weights are selected to minimize the sum of the squared e score values for the sample as a whole.

Because the variables may not all be based on the same scale, i.e., the variables may have different standard deviations, and because the standard deviations are usually considered arbitrary (e.g., the difference between the SD of a predictor involviong height in inches versus the SD of a predictor for the same sample measured as height in centimeters is usually considered not to be meaningful), the b weights are usually standardized prior to interpretation. As an indication of the relative importance of each of the variables in the equation, the researcher will usually examine the relative magnitude of the standardized b weights, known as the beta weights (Cooley & Lohnes, 1971; Pedhazur, 1982).

Many researchers assume that the magnitude of these weights can be compared to determine which variables are the most useful predictors; that is, which variable, or set of variables, accounts for the most variance in the dependent measure, Y (Cooley & Lohnes, 1971). It should be noted, however, that the beta weight of a variable reflects not only the effect of that variable, but also the covariances of the variable with all the other predictors included in the model (Pedhazur, 1982). Thus, the interpretation of beta weights to evaluate predictor variable importance is valid only when the predictor variables are all perfectly uncorrelated. When this is the case, the beta weights for each predictor variable, X_i, is equivalent to the correlation coefficient between the predictor variable and Y, $r_{y \cdot x_i}$.

The square of $r_{y \cdot x_i}$ is a measure of how much variance in Y is accounted for by the predictor variable, X_i.

But it is very rare for all of the predictor variables to be perfectly uncorrelated, particularly, in the behavioral and social sciences. When the predictor variables are correlated with each other, a condition referred to as collinearity (or multicollinearity) exists (Pedhazur, 1982). Researchers have noted three potentially noteworthy harmful effects which can occur as a result of multicollinearity.

First, the computational accuracy of least-squares calculations can be affected in extreme cases of collinearity (Belsley, Kuh, & Welsh, 1980). Second, as Pedhazur (1982) indicates, the statistical accuracy of test statistics can be affected by collinearity. And third, the accuracy of the interpretation of one's results can be affected by collinearity. As explained by Pedhazur (1982):

> It should be clear now that high multicollinearity may lead not only to serious distortions in the estimations of regression coefficients but also to reversals in their signs. Therefore, the presence of high collinearity poses serious threats to the interpretation of the regression coefficients [beta weights] as indices of effects. (p. 246)

Collinearity can exists for a variety of reasons. As pointed out by Belsley and his associates (1980), the existence of collinearity can simply be a reflection of the reality of that is being investigated by the researcher. Thompson and Borrello (1985) note that researchers often introduce multicollinearity into a study on purpose. This can be done for theoretical reasons, or because a researcher wishes to "multioperationalize" important variables or variables that are particularly difficult to measure. Whatever the reason, when the predictor variables in a regression analysis are intercorrelated, collinearity exists and it is no longer appropriate to interpret the magnitude of the beta weights as the sole indicator of a variable's predictive power.

The structure coefficient of a predictor variable, however, is not influenced by collinearity and can, as many researchers have suggested, be utilized in the interpretation of a regression equation to help evaluate the contribution of each of the predictor variables (Cooley & Lohnes, 1971; Kerlinger & Pedhazur, 1973; Levine, 1977; Thompson & Borrello, 1985). Thompson and Borrello (1985) have defined a structure coefficient as "the correlation between a predictor variable and a composite derived by weighting and aggregating the criterion variable(s)" (p. 205). In terms of regression, a structure coefficient is simply the correlation between a predictor variable, X_i, and the predicted score, Y*, that results from the prediction equation. As a measure of the degree of linear relationship between a predictor variable and the predicted score, a structure coefficient is not inflated or surpressed by collinearity (Thompson & Borrello, 1985). Therefore, the magnitudes of the structure coefficients of a set of predictor variables can be compared to assess the relative contribution of each. Furthermore, because a structure coefficient is a correlation coefficient, these coefficients can be squared to determine how much of the variance in the predicted scores, Y*, can be explained by each predictor. At this point a small heuristic example may be useful to illustrate the concept of a structure coefficient and how its interpretation compares to an interpretation based on the regression coefficients, i.e., the beta weights.

A Heuristic Example

A multiple regression analysis with three independent variables was conducted on the data set presented in Table 1. From the correlation matrix for these data, presented in Table 2, it can be seen that all three predictors appear to be correlated with Y to varying degrees, with X_2 having the largest correlation coefficient, $r_{y \cdot x2} = 0.7792$. Further examination of the intercorrelations between the predictors also indicates that there is some degree of collinearity present.

Insert Tables 1 and 2 about here.

The beta weights and structure coefficients for each of the three predictor variables are presented in Table 3. An examination of the beta weights indicates that of the three variables, X_2 is the best predictor, with a beta weight of $b_2 = 0.6660$, and appears to contribute the most to the regression equation. One would also conclude that with a beta weight of $b_3 = 0.0238$, X_3 does not appear to contribute to the regression equation much at all. Although the order of importance of the predictor variables does not change when the structure coefficients are interpreted instead of the beta weights, it can be seen that X_3 does in fact potentially contribute to the regression equation, $r_{y^*,x_3} = 0.4169$. Furthermore, the structure coefficient for X_1 is more than twice the magnitude of its beta weight. Thus, an interpretation based on the structure coefficients would indicate that while X_2 is the best overall predictor of the three variables, accounting for slightly more than 88% ($r_s^2 = .9388^2$) of the variance in Y*, both X_1 and X_3 also contribute or potentially contribute to the regression equation, each accounting for 42% and 17% of the variance in Y*, as indicated by squaring their structure coefficients reported in Table 3. An interpretation based on the beta coefficients only would have led to the conclusion that X_3 should be disregarded as a predictor of Y.

Beta weights can be used to address the question, "Assuming the set of predictors as a fixed given, how much does each predictor contribute to the prediction, disallowing double (or triple, or quadruple, etc.) counting for shared predicted variance originating in any two or more predictors?" Structure coefficients can be used to address the question, "Assuming the predicted portion of Y, Y*, as a fixed given, how useful would each predictor taken alone be at reproducing Y*?" Of course, since each beta weight equals each $r_{y \cdot x_i}$ when the predictors are uncorrelated, and since each structure coeficient always

equals $r_{y \cdot x_i}$ / R, when the predictors are uncorrelated beta and structure coefficients will yield identical interpretations regarding predictor variable importance.

Insert Table 3 about here.

Structure Coefficients and Other Parametric Procedures

As it was stated previously, it is a recognized fact that all parametric procedures can be derived from the same general linear model. As such, Thompson (1988) notes that every parametric procedure involves the creation of a "synthetic" score(s) (e.g., Y*) for each individual on some latent construct. This synthetic score is the composite score referred to above in the definition of a structure coefficient that was "derived by weighting and aggregating the criterion variable(s)". In regression analysis, the synthetic variable is the predicted dependent variable, Y*. The weights used to create this score are the beta weights. In factor analysis the factor scores for each person are the synthetic scores created from weights referred to as the factor pattern coefficients (Gorsuch, 1983; Thompson, 1988). Similarly, in discriminant analysis, the synthetic scores are the discriminant scores created with the discriminant function coefficients, and in canonical correlation the canonical variates are created using the standardized canonical function coefficients (Pedhazur, 1982; Thompson, 1988). Thus, it can be seen that for every parametric procedure there is a set of synthetic variables, analogous to the predicted score in regression, that is calculated for the variables and a set of least squares weights, analogous to the regression coefficients, i.e., beta weights (Thompson, 1988). Furthermore, for every procedure, a structure coefficient is simply the correlation between synthetic variable

scores and the set of variables employed in calculating the weights that are used to create the synthetic variable scores.

The importance of interpreting structure coefficients as opposed to canonical function coefficients in canonical analysis was emphasized long ago by Meredith (1964), who stated that "if the variables within a each set are moderately intercorrelated the possibility of interpreting the canonical variates by inspection of the appropriate regression weights [function coefficients] is practically nil" (p. 55). Thus, researchers have long recognized that collinearity can lead to misinterpretation of results and have continued to suggest that structure coefficients be employed as a remedy (Cooley & Lohnes, 1971; Gorsuch, 1983; Kerlinger & Pedhazur, 1973; Thompson & Borrello, 1985). As was demonstrated in the previous example presented in Table 3, interpretation of the standardized regression coefficients (beta weights) as compared to the structure coefficients can often lead to very different conclusions. Thus, if researchers wish to avoid serious misinterpretations that can result from interpreting only standardized weights, regardless of the procedure being employed, it is essential that structure coefficients also be calculated and interpreted.

It is important to note that regression coefficients and structure coefficients represent two different types of information. The regression coefficient is a measure the unique contribution of that particular variable to the latent construct under consideration over and above that of the other variables in the equation. The structure coefficient, on the other hand, represents the total contribution of that variable without taking into consideration any of the other variables in the model (Haase & Ellis, 1987). This is true for all parametric procedures. The researcher should be clear as to what type of information is of interest and be sure to utilize the appropriate coefficient in the interpretation of his/her results. However, when forced to report a single coefficient, Thompson (1988) argues that in the canonical case, the function coefficients might be ignored in favor of the structure coefficients. Thompson (1988) states:

> Since structure coefficients inform the researcher of the correlation between each variable and the synthetic variables, these coefficients are what inform the researcher regarding the meaning of what is actually being correlated in a given analysis. (pp. 15-16)

It has been noted before that all parametric procedures are "special cases" of canonical correlation. Thus, the argument favoring structure coefficients over their regression coefficient counterparts can readily be extended to all parametric procedures.

References

Belsley, D. A., Kuh, E., & Welsh, R. E. (1980). Regression diagnostics: Identifying influential data and sources of collinearity. New York: John Wiley & Sons.

Baggaley, A. R. (1981). Multivariate analysis: An introduction for consumers of behavioral research. Evaluation Review, 5, 123-131.

Cohen, J. (1968). Multiple regression as a general data-analytic system. Psychological Bulletin, 70, 426-443.

Cooley, W. W., & Lohnes, P. R. (1971). Multivariate data analysis. New York: John Wiley & Sons.

Daniel, L. G. (1990, January). Use of structure coefficients in multivariate education research: A heuristic example. Paper presented at the annual meeting of the Southwest Educational Research Association, Austin. (ERIC Document Reproduction Service No. ED 315 451)

Edwards, A. L. (1985). Multiple regression and the analysis of variance and covariance (2nd ed.). New York: W. H. Freeman.

Fornell, C. (Ed.) (1982). A second generation of multivariate analysis (Vol. 2). New York: Praeger.

Gorsuch, R. L. (1983). Factor analysis (2nd ed.). Hillsdale: Lawerence Erlbaum.

Haase, R. F., & Ellis, M. V. (1987). Multivariate analysis of variance. Journal of Counseling Psychology, 34, 404-413.

Kerlinger, F. N., & Pedhazur, E. J. (1973). Multiple regression in behavioral research. New York: Holt, Rinehart, & Winston.

Knapp, T. R. (1978). Canonical correlation analysis: A general parametric significance testing system. Psychological Bulletin, 85, 410-416.

Meredith, W. (1964). Canonical correlations with fallible data. Psychometrika, 29, 55-65.

Pedhazur, E. J. (1982). Multiple regression in behavioral research (2nd ed.). New York: Holt, Rinehart, & Winston.

Perry, L. (1990, January). The use of structure coefficients in regression research. Paper presented at the annual meeting of the Southwest Educational Research Association, Austin. (ERIC Document Reproduction Service No. ED 315 448)

Thompson, B. (1980, April). Canonical correlation: Recent extensions for modeling educational processes. Paper presented at the annual meeting of the American Research Association, Boston. (ERIC Document Reproduction Service No. ED 199 269)

Thompson, B. (1984). Canonical correlation analysis: Uses and interpretation. Newbury Park: SAGE.

Thompson, B. (1988, April). Canonical correlation analysis: An explanation with comments on correct practice. Paper presented at the annual meeting of the American Research Association, New Orleans. (ERIC Document Reproduction Service No. ED 295 957)

Thompson, B., & Borrello, G. M. (1985) The importance of structure coefficients in regression research. Educational and Psychological Measurement, 45, 203-209.

Willson, V. L. (1980). Research techniques in AERJ articles: 1969 to 1975. Educational Researcher, 9, 5-10.

Table 1
Data set for Regression Analysis with Predicted Values and Residuals

X1	X2	X3	Y	Predicted Value	Residual	X1	X2	X3	Y	Predicted Value	Residual
11	38	10	15	16.8309	-1.8309	14	24	12	15	14.6469	0.3531
07	42	16	18	16.6341	1.3659	10	26	12	14	13.8274	0.1726
12	38	18	17	17.3714	-0.3714	09	18	14	13	11.7371	1.2696
13	36	15	16	17.1507	-1.1507	11	30	16	16	15.1704	0.8296
14	40	15	17	18.3831	-1.3831	09	26	13	13	13.5360	-0.5360
15	32	11	14	16.7648	-2.7648	07	18	11	11	11.0160	-0.0160
05	20	13	10	10.8899	-0.8899	10	10	17	06	10.3126	-4.3126
14	44	18	21	19.3794	1.6206	09	12	08	12	10.2011	1.7989
14	34	12	17	16.9301	0.0699	10	32	18	18	15.3633	2.6367
10	28	16	11	14.3947	-3.3947	10	18	14	12	12.0562	-0.0562
08	24	10	13	12.6773	0.3227	16	20	15	16	14.4547	1.5453
16	30	16	18	16.7656	1.2344	10	18	12	15	12.0008	2.9992
15	26	15	16	15.5056	0.4944						

Note: From Edwards (1985).

Table 2
Correlation Matrix

	Y	X1	X2	X3
X1	0.5372			
X2	0.7792	0.3461		
X3	0.3462	0.2396	0.3756	

Table 3
Regression Results

Variable	Regression Beta Weights	Structure Coefficients
X1	0.3010	0.6473
X2	0.6660	0.9388
X3	0.0238	0.4169

Appendix A:
SAS Program to Analyze Hypothetical Data from Table 1

```
DATA D1; INFILE ABC;
INPUT ID X1 X2 X3 Y;
PROC PRINT;
PROC REG;
MODEL Y=X1 X2 X3/P STB;
OUTPUT OUT=D3 P=YHAT R=E;
PROC CORR;
 VAR Y X1 X2 X3 YHAT E;
```